Erschienen in: MUTHUKRISHNAN, Muthu, ed. and others. COSN '13 : Proceedings of the First ACM Conference on Online Social Networks. New York, NY: ACM, 2013, pp. 39-50. ISBN 978-1-4503-2084-9
https://doi.org/10.1145/2512938.2512946

Social Resilience in Online Communities: the Autopsy of Friendster

David Garcia, Pavlin Mavrodiev, and Frank Schweitzer
Chair of Systems Design, ETH Zurich
Weinbergstrasse 56/58, 8092 Zurich, Switzerland
dgarcia@ethz.ch, pmavrodiev@ethz.ch, fschweitzer@ethz.ch

ABSTRACT

We empirically analyze five online communities: Friendster, Livejournal, Facebook, Orkut, and Myspace, to study how social networks decline. We define social resilience as the ability of a community to withstand changes. We do not argue about the cause of such changes, but concentrate on their impact. Changes may cause users to leave, which may trigger further leaves of others who lost connection to their friends. This may lead to cascades of users leaving. A social network is said to be resilient if the size of such cascades can be limited. To quantify resilience, we use the k-core analysis, to identify subsets of the network in which all users have at least k friends. These connections generate benefits (b) for each user, which have to outweigh the costs (c) of being a member of the network. If this difference is not positive, users leave. After all cascades, the remaining network is the k-core of the original network determined by the cost-to-benefit (c/b) ratio. By analysing the cumulative distribution of k-cores we are able to calculate the number of users remaining in each community. This allows us to infer the impact of the c/b ratio on the resilience of these online communities. We find that the different online communities have different k-core distributions. Consequently, similar changes in the c/b ratio have a different impact on the amount of active users. Further, our resilience analysis shows that the topology of a social network alone cannot explain its success of failure. As a case study, we focus on the evolution of Friendster. We identify time periods when new users entering the network observed an insufficient c/b ratio. This measure can be seen as a precursor of the later collapse of the community. Our analysis can be applied to estimate the impact of changes in the user interface, which may temporarily increase the c/b ratio, thus posing a threat for the community to shrink, or even to collapse.

Keywords

game theory; social network analysis; social resilience; rational behaviour

1. INTRODUCTION

Online Social Networks (OSN), such as `Facebook` or `Friendster`, can quickly become popular, but can also suddenly lose large amounts of users. The appearance of competing OSN, with different functionalities and designs, create unexpected shifts of users that abandon one community for another [15]. While the dynamics of growth in these online communities are an established research subject [3, 21], there are still many open questions regarding the decline of online communities, in particular related to large OSN [37]. What are the reasons behind the decision of users to stop using an OSN? What is the role of the social network in keeping user engagement, or in the spreading of user dissatisfaction? Are there network structures that lead to higher risks of massive user departures? In this article, we assess the question of the relation between the topology of the user network, and the cascades of user departures that threaten the integrity of an online community. We build on previous theoretical work on network effects [5], providing the first empirical study of this phenomenon across successful, failed, and declining OSN.

The most successful OSN attract millions of users, whose interactions create emergent phenomena that cannot be reduced back to the behavior of individual users. The OSN is a communication medium that connects a large amount of people, which would stay together only if their interaction dynamics leads to the emergent entity that we call *the community*. The OSN and its users form a socio-technical system in which the persistence of the community depends on both the social interaction between users, and the implementation and design of the OSN. In this context, the *social resilience* [1] of an online community is defined as *"The ability of the community to withstand external stresses and disturbances created by environmental changes"*. In particular, the technological component of the OSN can change the environment of the users, and create stress that threatens the cohesion of the community. As an example, changes in the user interface pose a general risk for user engagement in OSN.

The fast pace of the Internet society has already led to the total disappearance of some very large online communities. The most paradigmatic example is `Friendster`, one of the first and largest OSN, which suffered a massive exodus of users towards competing sites. This led to its closure in 2011, to reopen as an online gaming without its profile data. As a reaction, the `Internet Archive` [1] crawled as much information as possible, creating a timeless snapshot of `Friendster` right before its closure. If, on the other hand, `Friendster` was still an alive and active community, this data would have been kept private and never made accessible at such scale. Before closure, users were warned and offered to delete their data from the site, leaving all the remaining data from this community as one of the largest publicly available datasets on social behavior.

The decay of `Friendster` is commented in a comedy video of the Onion News, in which a fictitious *"Internet archaelogist"* explains `Friendster` as an ancient civilization [2]. While proposed as a satire of the speed of Internet culture, this video illustrates the opportunities that a failed OSN offers for research. The users of such a community leave traces that allow us to investigate its failure. In this sense, we can name our work as *Internet Archeology*, because we analyze non-written traces of a disappeared society, aiming at understanding the way it worked and the reasons for its demise.

In this paper, we provide a quantitative approach to the collective departure of users from OSN. We start from a theoretical perspective that, under the assumption of rational user behavior, allows us to define a new metric for the relation between network topology and massive user leaves. We apply this metric to high quality datasets from `Friendster` and `Livejournal`, comparing their social resilience with partial datasets from `Facebook`, `Orkut`, and `Myspace`. The research presented here is based on publicly available datasets, allowing the independent validation of our results, as well as the extension to further analyses [19]. We find that social resilience differs greatly across the different networks we study. Interestingly, however, more resilient networks are not necessarily more successful. This indicates that success and failure cannot be explained by topology alone. Instead, envrironmental factors, e.g. competition, design choices, user behaviour, etc., play a considerable role in the faith of an OSN. As an application of our analysis, we focus on the time evolution of `Friendster`, tracking the changes in its social resilience and investigating how it decayed to a complete collapse. We finish by commenting on the limitations and extensions of our approach, and outline possible future applications.

2. RELATED WORK

Recent research has focused on the question of growth and decay of activity or interest-based social groups [24]. This line of research analyzes social groups as subcommunities of a larger community, tied together due to underlying common features of their members. Such approach can be equally applied to scientific communities and online social networks [3, 38, 34], revealing patterns of diffusion and homophily that respectively spread group adoption, and increase internal connectivity. In particular, the big datasets provided by online communities allow the study of group creation and maintenance [21]. These results lead to applied techniques to predict the fate of interest-based groups, and to improve clustering analysis of social networks. Our work differs from these previous results in the scope of our analysis: Instead of looking at small to medium sized groups within larger communities, we look at the OSN as a whole. In our approach, users are not connected to each other due to certain common interest or affiliation, but through an online platform that maintains their social links and serves as communication medium.

Another research topic close to our work is the analysis of individual churn, defined as the decision of a user to stop using a service in favor of a competitor. This topic has received significant attention due to its business applications, where previous works explore how individual users disconnect from P2P networks[18], and stop using massive multiplayer online games [22]. Regarding OSN, a recent study shows the relation between social interaction and user departure in the online community `Yahoo answers` [10]. Furthermore, the same question has been addressed in a recent article [37], analyzing a mysterious online social network of which nor the name, size, nor purpose is explained. While these results are relevant for the question of user engagement, it is difficult to consider them in further research if we do not have information about the nature of the studied network. Social networks can have very different roles in online communities, requiring a differentiation between traditional social networking sites and online communities with a social network component, but where social interaction is mediated through other channels. The results of [37] reveal that 65% of the users that have no friends still remain active after three months, indicating that such social network is not precisely necessary for a user to use the site. As an example, a `Youtube` user does not need to create and maintain social contacts to interact with other users, which can be done through videos and comments independently of the social network.

Our work complements the previous results on individual user departures mentioned above, as we analyze the social resilience of the online community at the collective level. We build on these empirically validated microscopic rules of churn, to focus on cascades of departures through large OSN. We analyze the macroscopic topology of the social network and its role in the survival of the community. This kind of macroscopic effects are relevant to study the emergence of social conventions [25], an dynamics of politically aligned communities [14], in addition to the case of OSN we address here.

The particular problem of enhancing resilience by fixing nodes of a social network has been proposed and theoretically analyzed [5], aiming to prevent the *unraveling* of a social network. This implies that social resilience can be analyzed through the k-core decomposition of the social network, as explained in Section 3.1 . In addition, k-core centrality is the current state-of-the art metric to find influential nodes in general networks [23]. Regarding OSN, the k-core decomposition was applied for a global network of instant messaging [26], as well as for the Korean OSN `Cyworld` [2, 8], motivated by user centrality analysis rather than social resilience. To our knowledge, this article introduces the first empirical analysis

[1] http://archive.org/details/friendster-dataset-201107
[2] https://www.youtube.com/watch?v=7mFJdOsjJ0k

of social resilience, relating changes in user environment with cascades of departing users, through analysis based on the the k-core decomposition of different OSN.

3. SOCIAL RESILIENCE IN OSN

3.1 Quantifying Social Resilience

A characteristic property of any online social network is the presence of influence among friends. In particular, individual decisions regarding participating or leaving the network are, to a large extent, determined by the number of one's friends and their own engagement [3]. Therefore, users leaving a community have negative indirect effects on their friends [37]. This may trigger the latter to also leave, resulting in further cascades of departing users which may ultimately endanger the whole community. Social resilience acts to limit the spread of such cascades.

One approach to quantify social resilience is by natural removal of nodes based on some local property, for example degree [26]. By studying the network connectivity after such removals, one can identify nodes with critical importance for keeping the community connected. Importantly, by focusing on local properties we can only quantify the direct effects that a node removal has on the connectivity of the network.

In this paper, we propose an extension based on the k-core decomposition [23]. A k-core of a network is a sub-network in which all nodes have a degree $\geq k$. The k-core decomposition is a procedure of finding all k-cores, $\forall k > 0$, by repeatedly pruning nodes with degrees k. Therefore, it captures not only the direct, but also the indirect impact of users leaving the network. As an illustration consider Figure 1, which shows targeted removal of nodes with degrees < 3. On one hand,

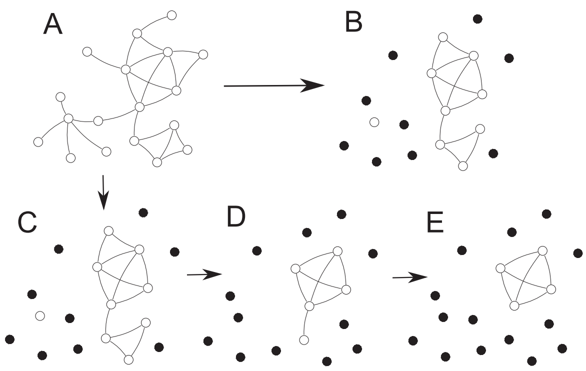

Figure 1: Effects of node removals on network connectivity as captured by degree only (A → B) and k-core decomposition (A → C → D → E)

starting from the network in A and removing all nodes with degrees < 3, produces the network in B. The black nodes in B have been removed (and thus are disconnected), and the final network consists of the 9 white nodes. The transition A → B shows only the direct effects of users with < 3 friends leaving.

On the other hand, starting again from A, and applying the k-core procedure, will repeatedly remove nodes until only those with degrees ≥ 3 remain. The first step, A → C, removes the same black nodes as before. Continuing, C → D, removes those nodes that have been left with < 3 neighbours in C, and disconnects them as well. The final step, D → E, finishes the process by disconnecting the last white node in D that was left with < 3 friends. As a result, the final network is the fully connected network of the 4 white nodes. Hence, supposing that users leave a community when they are left with less than 3 friends, the k-core decomposition captures the full cascading effect that departing users have on the network as a whole.

We proceed by formalizing social resilience based on a *generalized* k-core decomposition. To this end, we present a theoretical model in which rational users decide simultaneously either to stay in the network or to leave it. These decisions are based on maximizing a utility function that weighs the benefits of membership against the associated costs. We show that the equilibrium network which maximizes the total payoff in the community, corresponds to a generalized k-core decomposition of the network.

3.2 Generalized k-core decomposition

Following [17], we extend the traditional k-core decomposition by recognizing that the pruning criterion need not be limited to degree only. Let us define a *property* function $\mathcal{B}_i(H)$ that given a sub-network $H \subseteq G$ associates a value, $n_i \in \mathbb{R}$, to node i. A generalized k-core of a network G is, then, defined as a sub-network $H \subseteq G$, such that $\mathcal{B}_i(H) \geq k$, $\forall i \in H$ and $k \in \mathbb{Z}$. The general form of B_i allows us to model different pruning mechanisms. For example, the traditional definition of the k-core can be recovered in the following way – for every node i take its immediate neighbourhood, $\mathcal{N}_i$, and fix $B_i(H) := |\mathcal{N}_i|$, $\forall H \subseteq G$. Other authors have also shown that considering weighted links in $\mathcal{B}_i$ can more accurately reveal nodes with higher spreading potential in weighted networks [13].

Note that by definition higher order cores are nested within lower order cores. We use this to define that a node i has *coreness* k_s if it is contained in a core of order k_s, but not in a core of order $k' > k_s$.

3.3 A rational model for OSN users

Here, we model the cost-benefit trade-off of OSN users in the following way. Assume that users in a given network, G, incur a constant integer cost, $c > 0$, for the effort they must spend to remain engaged. Accordingly, they receive a benefit or payoff from their friends in the network. Let the benefit of user i be the property function $\mathcal{B}_i(H)$ with $i \in H$. Assume non-increasing marginal benefits with respect to the size of H, i.e. $\mathcal{B}_i''(H) \leq 0$, otherwise costs are irrelevant as any cost level could be trivially overcome by increasing the size of H. This assumption is also supported by other empirical investigations of large social networks which show that the probability of a user to leave is concave with the number of friends who left [3, 37].

Users can select one of two possible actions – `stay` or `leave`. The utility of user i, is $U_i = 0$, if he chose `leave` or $U_i = \mathcal{B}_i(H) - c$, for `stay`. Finally, since users are rational, they will try to maximize their utility and so will choose `stay` as long as $U_i > 0$.

It is easily seen that the equilibrium network, G^*, which maximizes the total utility, $U(G) = \sum_i U_i$, is composed of users who choose `stay` when $c < k_s^i$, and `leave` otherwise. In other words, node i should remain engaged in the network as long as the cost, c, does not exceed its generalized coreness, k_s^i. In this sense, G^* corresponds to the generalized k-core of G.

To illustrate that G^* is indeed an equilibrium network, we need to show that no user has an incentive to unilaterally join it or leave it. Consider a node, $j \in G^*$ who chooses `stay`. This node would belong to a generalized k-core, k_s^j, and by definition, $B_j(H) - k_s^j \geq 0$. Since, j stayed in the network, it must be that $c < k_s^j$, therefore $B_j(H) - c > 0$. So, j will be forfeiting positive utility, should he decide to leave. In the same manner, consider another node $l \notin G^*$ who chooses `leave`, thus his coreness $k_s^l \leq c$. All his friends with the same coreness would have left the network, therefore the only benefit that l could obtain from staying would come from his connections with nodes in higher cores. The benefit, B_l, from such connections must not exceed k_s^l, otherwise l would have belonged to a higher core in the first place. Since $k_s^l \leq c$ we have $B_l < c$. This implies that l necessarily obtains negative utility from staying, so he has no incentives to do so. Moreover, G^* is optimal, as we showed that any change from the equilibrium actions of any user inevitably lowers his utility and decreases the total utility in the network. We also argue that it is reasonable to expect this equilibrium network to be reached in an actual setting, since it maximizes the utility of all users simultaneously, as well as the welfare of the network provider.

In the rest of the paper, we approximate B_i as proportional to the number of i's direct friends, N_i, i.e. $B_i = bN_i$, for some $b \in \mathbb{Z}$. Taking k_s^i to be the coreness of i, by definition it holds that $bN_i \geq k_s^i$. The maximum cost, c, that i would tolerate as a member of the community must be strictly smaller than its coreness, hence $bN_i > c$ and $N_i > c/b$. The last result implies that the minimum number of friends that a node i needs to remain engaged must be strictly larger than c/b. Therefore, $k_s^i \geq K$, i.e. the coreness of a participating user i must not fall below a critical value K with K given by:

$$K = (c/b) + 1 \tag{1}$$

Based on the above discussion, we see that a user will remain in a network with a high c/b ratio if its coreness k_s is high. This is because, by definition, i is part of a connected network of nodes with large minimum degrees and hence large benefits. In contrast, simply having a large degree does not imply that a user will obtain large utility from staying. Note that a high-degree node may nevertheless have low coreness. This means that i would be part of a sub-network in which all nodes have low minimum degrees. As a result a lower c/b ratio would suffice to start a cascade of users departing, that can quickly leave i with no friends and thus drive it to leave too. Finally, we define social resilience of a community as the size of the K core. In other words, this is the size of the network that remains after all users with $k_s \leq c/b$ have been forced out. This definition allows us to quantify social resilience and reliably compare it across communities even for unknown c/b ratios, as shown in Section 5.

4. DATA ON ONLINE SOCIAL NETWORKS

For our empirical study of social network resilience, we use datasets from five different OSN. The choice of these datasets aims at spanning a variety of success stories across OSN, including successful and failed communities, as well as communities currently in decline. The size, data gathering methods, and references are summarized in Table 1, and outlined in the following.

Friendster
The most recent dataset we take into account is the one retrieved by the Internet Archive, with the purpose of preserving `Friendster`'s information before its discontinuation. This dataset provides a high-quality snapshot of the large amount of user information that was publicly available on the site, including friend lists and interest-based groups. In this article, we provide the first analysis of the social network topology of `Friendster` as a whole.

Since some user profiles in `Friendster` were private, this dataset does not include their connections. However, these private users would be listed as contacts in the list of their friends who were not private. We symmetrized the `Friendster` dataset by adding these additional links. Due to the large size of the `Friendster` dataset, we symmetrized the data by using Hadoop, which we distribute under a Creative Commons license [3].

Livejournal
In `Livejournal`, users keep personal blogs and define different types of friendship links. The information retrieval method for the creation of this dataset combined user id sampling with neighborhood exploration [29], covering more than 95% of the whole community. We choose this `Livejournal` dataset for its overall quality, as it provides a view of practically the whole OSN.

Note that the design of `Livejournal` as an OSN deviates from the other four communities analyzed here. First, `Livejournal` is a blog community, in which the social network functionality plays a secondary role. Second, `Livejournal` social links are directed, in the sense that one user can be friend of another without being friended back. In our analysis, we only include reciprocal links, referring to previous research on its k-core decomposition [23]. By including this dataset, we aim at comparing how different interaction mechanisms and platform designs influence social resilience.

Orkut
Among declining social networking sites, we include a partial dataset on `Orkut` [29], which was estimated to cover 11.3% of the whole community. Far from the quality of the two previous datasets, we include `Orkut` in our analysis due to its platform design, as this dataset includes users that did not have a limit on their amount of friends. Furthermore, `Orkut` has a story of local success in Brazil[4], losing popularity against other sites at the time of writing of this article.

MySpace
One of the most famous OSN in decline is `Myspace`, which was the leading OSN before `Facebook`'s success [15]. We include a relatively small dataset of 100000 users of MySpace [2], which was aimed to sample its degree distribution. This dataset was crawled through a Breadth-First Search method, providing a partial and possibly biased dataset of `Myspace`. We include

[3] web.sg.ethz.ch/users/dgarcia/Friendster-sim.tar.bz2
[4] http://www.digitaltrends.com/computing/facebook-taking-over-globally-with-almost-700-million-users/

Table 1: Outline of OSN and datasets

name	launch date	status in 2013	crawl date	users	links	source
Livejournal	1999	in decline	Dec 2006	5.2×10^6	2.8×10^7	[29]
Friendster	2002	discontinued	Jul 2011	1.17×10^8	2.58×10^9	Internet Archive
Myspace	2003	in decline	Oct 2006	10^5	6.8×10^6	[2]
Orkut	2004	in decline	Nov 2006	3×10^6	2.23×10^8	[29]
Facebook	2004	successful	May 2008	3×10^6	2.36×10^7	[36]

this dataset as an exercise to study the influence of sampling biases in the analysis of social resilience.

Facebook
We want to complete the spectrum of success of OSN, from the collapse of `Friendster` to the big success of `Facebook`. The last dataset we include is a special crawl which aims at an unbiased, yet partial dataset as close as possible to the whole community [36]. This dataset was retrieved based on regional networks, for which social connections among the members of that subnetwork were accessible at the time of the crawl.

The partial datasets on `Orkut`, `MySpace`, and `Facebook` allow us to analyze of OSN that are still "alive", in the sense that they have not been discontinued yet. As an analogy to the *autopsy* of `Friendster`, we provide a *biopsy* of the other OSN, taking a small sample due to their privacy and data availability issues. Our results on these datasets are valid to the extent of their publicly available data, while we can be confident that our analysis of `Livejournal` and `Friendster` are representative of their complete user bases.

5. EMPIRICS OF OSN RESILIENCE

5.1 K-core decomposition

Following the analysis of the model presented in Section 3.3, we computed the k-core decomposition for each of the OSN datasets introduced above. Among those datasets, `Friendster` and `Livejournal` cover the vast majority of their respective communities. Figure 2 shows a schematic representation of the k-core decomposition of `Friendster` and `Livejournal`. Each layer of the circles corresponds to the nodes with coreness k_s, with an area proportional to the amount of nodes with that coreness value. The color of each layer ranges from light blue for $k_s = 1$, to red for $k_s = 304$. The distribution of colors reveals a qualitative difference between both communities: `Friendster` has many more nodes of high coreness than `Livejournal`, which has a similar color range but many more nodes with low k_s. This difference indicates that, to keep together as a community, `Livejournal` needs to have a much lower c/b than `Friendster`. This scenario is rather realistic, as `Livejournal` is a blog community in which users create large amounts of original content.

Our theoretical argumentation, presented in Section 3.3, implies that node coreness is a more reasonable estimator for resilience than node degree. A degree of at least k_s is a necessary condition for a coreness of k_s, but a high degree does not necessarily mean a high coreness. Taking `Friendster` an example, Figure 3 shows the boxplot for the distribution of k_s versus node degree, indicating the spread of k_s for nodes of similar degree. The empirical data shows that a high degree does not necessarily mean a high k_s, even finding nodes with very low k_s and very high degree. Nevertheless, it is clear

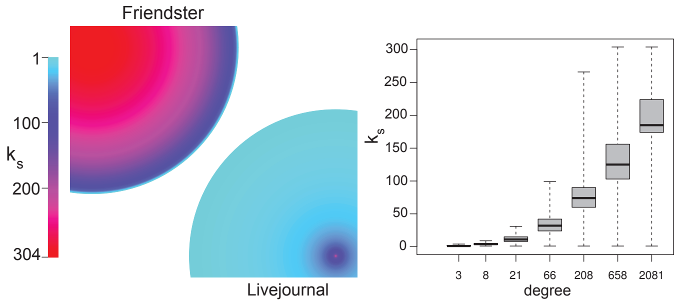

Figure 2: Left: Overview of the k-core decomposition for Friendster and Livejournal. Layers are colored according to k_s, with areas proportional to the amount of nodes with such k_s. Right: boxplot of k-shell indices by degree for `Friendster`. Dark lines represent the mean, and dashed bars show extreme values. Boxes are arranged in the x-axis according to the middle value of their bin.

that k_s is likely to increase with degree, but mapping degree to coreness would wrongly estimate the resilience of the community as a whole. By measuring coreness, we can detect that some nodes belong to the fringe despite their high degree, as the coreness integrates global information about the centrality of the node.

5.2 Resilience comparison

Extending the above observations, we computed the k-core decomposition of the three additional OSN, aiming at comparing their relation between their environment, measured through c/b, and the amount of users expected to be active under such conditions.

We focus our analysis on the Complementary Cumulative Density Function (CCDF) of each network, defined as $P(k_s > K)$. As shown in Section 3.3, the cost-benefit-ratio c/b corresponds to a value K that determines the nodes that leave the network, which are those k_s coreness below K (Eq. 1). Under this conditions, the CCDF of k_s measures the amount of nodes that will remain in the network under a given c/b, allowing us to compare how each OSN would withstand the same values of cost and benefit.

The right panel of Figure 3 shows the log-log CCDF of the five OSN. The first two communities to compare are `Livejournal` and `Friendster`, as the datasets on these two are the most reliable. First, the CCDF of `Friendster` is always above the CCDF of `Livejournal`. This is consistent with the structure shown in Figure 2, where it can be appreciated that `Livejournal` has many more nodes in the fringe than `Friendster`. Second, both CCDF reach comparable maximum values, re-

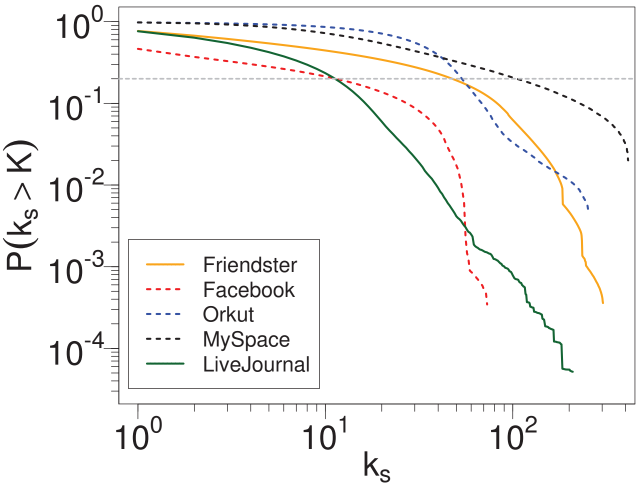

Figure 3: CCDF of k_s for all five OSN. The horizontal dashed line shows the cut at 0.2.

gardless of the fact that Friendster was 20 times larger than Livejournal. Such skewness in the coreness of Livejournal can be interpreted as a result of a higher competition for attention, as expected from a blog community in comparison with a pure social networking site, like Friendster was.

Focusing on the tails of the distributions, we can compare the patterns of resilience for environments with high K. The comparison between the resilience of these communities is heavily dependent of the value of K, as for example, Livejournal is less resilient than Facebook for values of K between 10 and 50, but more resilient below and above such interval. A similar case can be seen between Friendster and Orkut, as their CCDFS cross at 60 and 200. Thus, Friendster would be more resilient than Orkut if K lies in that interval, while Orkut would have a larger fraction of active nodes if $K < 60$ or $K > 200$.

It is important that these comparisons are made between the reliable datasets of Friendster and Livejournal, compared with partial datasets from the other communities. While our conclusions on the first two OSN can be seen as global findings on the community as a whole, the rest are limited to the size of the datasets available. A particularly clear example of the effect of the crawling bias is the distribution of coreness for Myspace, which shows an extreme resilience in comparison to all the other datasets, with the exception of Orkut for $K < 50$. As commented in Section 4 , the method used for Myspace was very biased towards nodes of high degree, leaving an unrealistic picture of the resilience of the whole community. Additionally, the low starting value of the CCDF of Facebook could be related to the crawling method of the dataset, restricted to regional networks. This highlights the importance of publicly available datasets for academic research: While we are able to make a major *autopsy* of Friendster and Livejournal, our analysis of the other three datasets can be considered a *biopsy*, as we can only use a small sample of them.

Regardless of any crawling bias, we found that these networks have maximum coreness numbers much higher than previous results. The maximum k_s found for the network of instant messaging was limited to 68 [26], and close to 100 for the OSN Cyworld [8]. Livejournal has a maximum k_s of 213, Friendster of 304, Orkut of 253, and Myspace as a very deep core of $k_s = 414$. The exception lies in the Facebook dataset, where we find a maximum k_s of 74. This evidence shows that OSN can have much tighter cores than the ones found in previous research, revealing that they contain small communities with very high resilience.

As a final comparison, we focus on the values of K for the catastrophic case of the networks losing 80% of their nodes, i.e. where the CCDF has a value of 0.2. The data shows that both Facebook and Livejournal would lose 80% of their users under a value of K close to 10. For the case of the unsuccessful communities of Orkut and Friendster, it requires a much worse environment, with values of K above 60. This way, the empirical data supports the idea that, under the same environmental conditions, Facebook and Livejournal are less resilient than the three other networks, which were less successful. This means that the topology of their social network is not enough to explain their collapse, but indicates that bad decisions in design and interface changes can spread through the network and drive many users away.

6. NOT POWER-LAW DEGREE DISTRIBUTIONS

In Section 3.3, we modelled the large-scale cascades of departures as the result of rational users evaluating their net benefits of staying in the network. However, investigating if OSN have power-law degree distributions is important, as it could provide an alternative model for user exodus. In particular, networks with power-law degree distributions do not have an epidemic threshold below which a "sickness" cannot spread [32]. Instead, the sickness will survive within the network for an unbound amount of time and eventually infect most of the nodes. Such sickness could be a meme or a social norm, but could also be the decision of leaving the community.

Power-law degree distributions arise from empirically tested mechanisms of preferential attachment [26], and bursty behavior in link creation [11, 30]. Numerous previous works have reported power-law degree distributions in social networks [2, 8, 26, 29]. Nevertheless, most of these works rely on goodness of fit statistics, and do not provide a clear test of the power-law hypothesis. It states that the degree distribution follows the equation $p(d) = \frac{\alpha-1}{\deg_{\min}}\left(\frac{d}{\deg_{\min}}\right)^{-\alpha}$ for $d \geq \deg_{\min}$. This is usually described as $p(d) \propto d^{-\alpha}$, and often argued as valid if metrics such as R^2, or F_1 are high enough. While a high goodness of fit could be sufficient for some practical applications, the power-law hypothesis can only be tested, and eventually rejected, through the result of a statistical test, assuming a reasonable confidence level.

We followed the state-of-the-art methodology to test power laws [9], which roughly involves the following steps. First, we created Maximum Likelihood (ML) estimators $\hat{\alpha}$ and $\widehat{\deg}_{\min}$ for $p(d)$. Second, we tested the empirical data above $\widehat{\deg}_{\min}$ against the power law hypothesis and we recorded the corresponding KS-statistics (D). Third, we repeated the KS test for 100 synthetic datasets that follow the fitted power law above $\widehat{\deg}_{\min}$. The p-value is then the fraction of the synthetic D values that are larger than the empirical one. Thus, for each

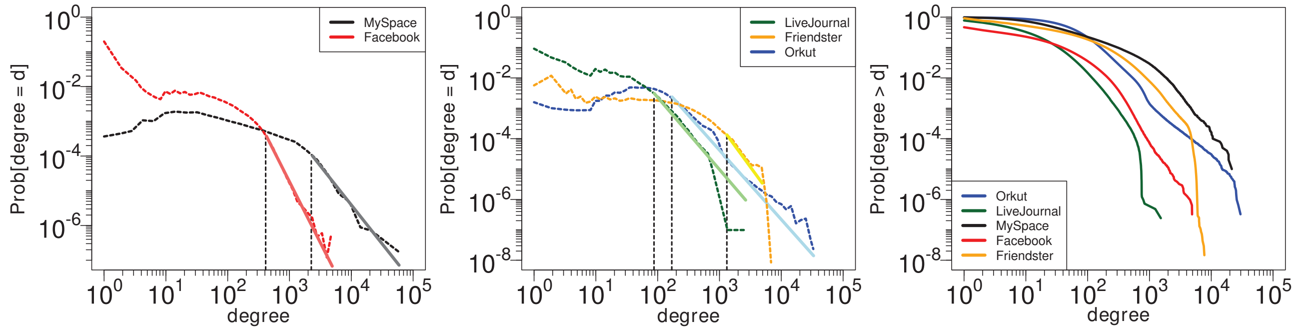

Figure 4: Complementary cumulative density function (cdf) and probability density functions (pdf) of node degree in the five considered communities. For each pdf, lighter lines show the ML power-law fits from $\widehat{\deg}_{\min}$. Vertical dotted lines indicate $\widehat{\deg}_{\min}$.

degree distribution, we have the ML estimates $\widehat{\deg}_{\min}$ and $\hat{\alpha}$, which define the best case in terms of the KS test, with an associated D value, and the p-value.

Ultimately, a power law hypothesis cannot be rejected if (i) the p-value of the KS-test is above a chosen significance level [9], and (ii) there is a sufficiently large amount of datapoints from $\deg_{\min}$ to $\deg_{\max}$ [33]. We found that the degree distributions of `Facebook`, `Friendster`, `Orkut` and `Livejournal` have p-values well below any reasonable significance threshold, showing an extremely reliable empirical support to reject the power-law hypothesis (Table 2).

Table 2: Power law fits of the degree distributions of the analyzed networks.

dataset	$\widehat{\deg}_{\min}$	$\hat{\alpha}$	n_{tail}	D	p
Friendster	1311	3.6	2.9×10^5	4.59	$< 10^{-15}$
LiveJournal	88	3.3	81141	0.02	$< 10^{-15}$
Facebook	423	4.6	4918	0.14	$< 10^{-15}$
Orkut	171	3	2.8×10^5	0.02	$< 10^{-15}$
MySpace	2350	3.6	623	0.03	0.22

For the case of `Myspace`, a KS test gives a p-value of 0.22, which can be considered high enough to not reject the power-law hypothesis [9]. Therefore `Myspace` satisfies the first criterion, but when looking at the range of values from $\deg_{\min}$ to $\deg_{\max}$ (roughly one order of magnitude), and the low amount of datapoints included, this KS-test composes a merely anecdotal evidence of the extreme tail of Myspace. If accepted, the power-law distribution would explain just 0.623% of the `Myspace` dataset. In addition, BFS methods have been shown to bias the macroscopic properties of the datasets they produce [31]. This leads to the conclusion that, while we cannot fully reject the power-law hypothesis, we can safely state that the dataset does not support the hypothesis otherwise. Figure 4 shows the degree distributions and their CCDF. For each OSN, we show how the typical log-log plot of the PDF is misleading, as a simple eye inspection would suggest power-law distributions, but a robust statistical analysis disproves this possibility.

7. THE TIME EVOLUTION OF FRIENDSTER

In this section, we describe a *post hoc* case study of the way how `Friendster` rose and collapsed, using the available timing information in the dataset.

7.1 Social growth mechanism

The `Friendster` dataset does not provide the date of creation of user accounts or social links, but it includes a user id that increased sequentially since the creation of the site. We analyzed the time series of `Friendster` in an event time scale, where each timestamp corresponds to the id of each user. We measured the time distance of an edge e, which connects users u_1 and u_2, as the difference between the ids of these users $d(e(u_1, u_2)) = |id(u_1) - id(u_2)|$. In the following, we show how early users connected to later users, making the network grow.

We divided the network in time slices of a width of 10 million user ids, with a last smaller slice of 7 million ids. Each of these 12 slices contains a set of nodes that have connections i) to nodes that joined the community before, ii) to nodes that joined the network afterwards, and iii) internally within the slice. This way, for the slice of time period t we can calculate its internal average degree $2|E_{in}(t)|/|N(t)|$, where $E_{in}(t)$ is the set of edges between nodes in the slice t, noted as $N(t)$.

As an extension, we define $E_p(t)$ and $E_f(t)$ as the sets of edges towards nodes that joined the community before t (past nodes), and nodes that joined after t (future nodes). We measured the time range of connections to the past $P(t)$ as the mean distance of the edges in $E_p(t)$, and the rage of connections to the future $F(t)$ as the mean distance of their future counterpart $E_f(t)$. By definition, the amount of past nodes for the first slice is 0, equally to the amount of future nodes for the last slice. If the process of edge creation was completely independent of these timestamps, the network would have some arbitrary sequence of node ids. In such network, $P(t)$ would steadily increase with each slice, having an expected value of $|N|/2$ for the last one, where $|N|$ is the size of the network. Similarly, $F(t)$ would decrease from $|N|/2$ at the first slice, converging at 0 in the last one.

The time evolution of the range of connections to past and future is shown in the left panel of Figure 5. Each circle represents a slice of the network, with growing t from left to right.

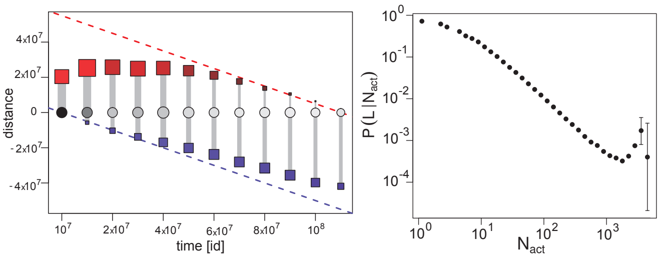

Figure 5: Left: Schema of connectivity of Friendster users across time. Each circle represents a slice of the network of width of 10 million user ids. Blue squares represent past users and red squares represent future users, with a distance from their slice according to $P(t)$ and $F(t)$ respectively. The dashed lines show the expectation of these two metrics in a random id sequence of the network. Right: Likelihood of a `Friendster` user to leave, given the amount of active friends of the user. The decreasing likelihood validates de assumption that users are more likely to leave when they do not have enough active friends.

Their horizontal alignment represents the present with respect to the slice, and each circle is connected to a blue square on the below that represents past nodes, and a red square above that represents future nodes. Circles have a size proportional to $|N(t)|$, which keep approximately constant throughout time. The darkness of each circle is proportional to its internal connectivity $|E_{in}(t)|$, and the width of the connections from circles to past and future squares are proportional to $|E_p(t)|$ and $|E_f(t)|$ respectively. Internal connectivities decrease through time, as early slices had significantly higher $|E_{in}(t)|$. This indicates that the initial root of users of `Friendster` was much more tightly connected among themselves than towards other nodes, creating a denser subcommunity of old users. A possible explanation for this pattern is that `Friendster` started as an OSN for dating, and its design was later shifted towards generalized networking as it became popular.

The squares of Figure 5 left are positioned according to the mean past $P(t)$ and future $F(t)$ distances of each slice. As a comparison with random network construction, dashed lines show their expected values as explained above. For early slices, the mean future distance is significantly lower than its random counterpart, revealing a connectivity pattern that limits the range of future connections. This shows a decay in the diffusion process through the offline social network, where the potential of a user to bring new users decreases through time. This suggests a possible "user expiration date" after which a user of a OSN cannot be expected to bring new users.

7.2 Microdynamics of user activity

We used the `Friendster` dataset to explore the empirical properties of the benefit function $\mathcal{B}_i(H)$, explained in Section 3.2. In our rational model, that function determines when a user i becomes inactive, given some quantifiable properties of its social environment H. While the dataset does not provide precise activity statistics to estimate $\mathcal{B}_i(H)$, we can estimate the conditions for users to become inactive through their sequence of ids. Following the methodology of [26], we approximate the time when a user became active as its id in the sequence, and the maximum timestamp associated to its edges as the time when it became inactive. This way, we only take into account user activity as link creation in the social network, leaving out other actions such as creation of messages or sharing of pictures.

For each user, we extracted the events when a friend becomes inactive, or when a friend joins the network, calculating the amount of active friends N_{act} of the user in each of those events. This value determines the period when the user is creating new friends, until N_{act} reaches its maximum value, after which the user has decreasing amounts of active friends and ends becoming inactive itself. After this maximum value, we calculate the likelihood of a user leaving the OSN (L) given its amount of active friends $P(L|N_{act})$, in order to provide a first estimation of the social conditions for users becoming inactive in `Friendster`. The right panel of Figure 5 shows this likelihood, revealing that users are much more likely to leave when they have low amounts of active friends. This validates our assumption that the benefits of a user are monotonically increasing with its amount of active neighbors, as they are much less likely to leave the OSN when they have a sufficient amount of active friends.

Two additional observations can be done about the likelihood $P(L|N_{act})$. First, the shape of its dependence of N_{act} reveals high variance, despite of its fast decrease. This indicates that the likelihood of users leaving scales with connectivity, i.e. the fraction of users likely to leave the OSN does not vanish when network size and density tend to infinity. Second, there is a small trend at the tail of the likelihood, where some values seem to increase. Our statistics do not fully validate the existence of this increase, as there are very few users with so many friends, but we can observe that the monotonically decreasing behavior up to that level does not exist any more. This suggests the presence of information overload [28], in which users with very large amounts of friends might be unable to cope with all the information provided by the OSN, and thus perceiving lower benefits.

7.3 Resilience and decline of Friendster

We combined the sequence of user ids with the k-core decomposition of `Friendster` to study how its resilience changed over time. In particular, we explored the relation between the coreness of users and the time when they joined the community. To analyze the changes in resilience, we divided the users along the median of the distribution of coreness values, $\bar{k_s} = 6$. This way, for each period of time, there is an amount of users in the lower half of the distribution ($k_s < \bar{k_s}$). When such amount increases, the new members that joined the OSN in that period are at higher risk to leave than when they have coreness values above the median. We measure the resilience of these time-dependent parts of network as the ratio between users with $k_s < \bar{k_s}$, and the total amount of users in the slice.

We created slices of 100000 user ids, calculating a point sample estimate of $P(k_s < \bar{k_s})$. Inset of Figure 6 shows the time evolution of this ratio, with a dark area showing 99% confidence intervals. First, we notice that the skewness of k_s does not affect our statistic, as the confidence intervals are sufficiently concentrated around the point estimates. Second, we can identify certain time periods when the new users of `Friendster` only connected to its fringe, having larger ratios of nodes with coreness below the median. The first moment with a peak is at the very beginning, to drop to ratios around 0.3 soon after. This shows that the set of very early users did not fully exploit the social network, and it took a bit of time for the OSN to become more resilient. The second peak is shortly after having 22 million users, which coincides with the decay of popularity of `Friendster` in the US. Finally, the ratio of users at risk went above 0.5 before the community had 80 million accounts, showing a lack of cohesion as its shutdown approaches, as new users do not manage to connect to the rest.

To conclude our analysis, we explored how the spread of departures captured in the k-core decomposition (see Section 3.3) can describe the collapse of `Friendster` as an OSN. As we do not have access to the precise amount of active users of `Friendster`, we proxy its value through the `Google` search volume of *www.friendster.com*. The inset of Figure 6 shows the relative weekly search volume from January 2009. At some point in 2009, `Friendster` introduced changes in its user interface, coinciding with some technical problems, and the rise of popularity of `Facebook` [5]. This led to the fast decrease of active users in the community, ending on its discontinuation in 2011.

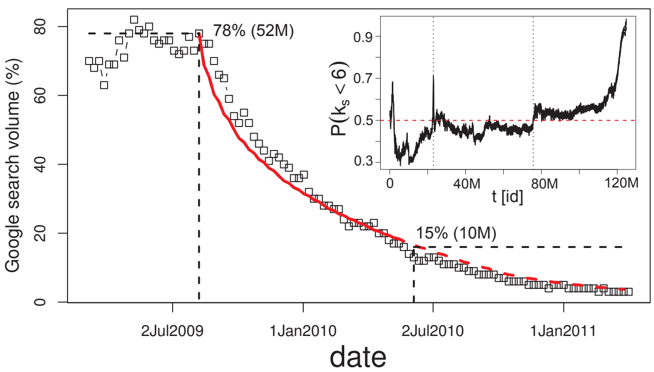

Figure 6: Weekly Google search trend volume for `Friendster`. The red line shows the estimation of the remaining users in a process of unraveling. Inset: time series of fraction of nodes with $k_s < 6$.

We scale the search volumes fixing 100% as the total amount of users with coreness above 0, 68 million. At the point when the collapse of `Friendster` started, the search volume indicates a popularity of 78% of its maximum. We take this point to start the simulation of a user departure cascade, with an initial amount of 58 million active users, i.e. users with coreness above 3. The second reference point we take is June 2010, when Friendster was reported to have 10 million active users [6], corresponding to 15% of the 68 million user reference explained above. The search volume on that date is 14%, showing the validity of the assumption that the maximum amount of active users corresponds to those with coreness above 0. Thus, these 10 million remaining users correspond to nodes with $k_s > 67$.

Given these two reference points, we can approximate the collapse of the network through its "unraveling" per k-core. Our assumption is that a critical coreness K_t, as defined in Eq.1, starts at 3 and increases by 1 at a constant rate. Such K_t is the result of an increasing cost-to-benefit ratio, and thus all the nodes with $k_s < K_t$ would leave the community. Then, for each timestep, the amount of remaining users would correspond to the CCDF shown in Figure 3. In this simulation, K increases at a rate of 6 per month, i.e. from 3 to 67 between our two reference points.

The red line of Figure 6 shows the remaining users under this process, with dashed values after the second reference point of June 2010. We can observe that this process approximates well the decay of `Friendster` from the start of its decline, to its discontinuation in 2011. The R^2 value for this fit is 0.972, leaving some slight underfit through 2009. This fit show the match between two approximations: on one side the search volume as an estimation of the amount of active users, and on the other side the amount of remaining users when the c/b ratio increases constantly through time.

[5]www.time.com/time/business/article/0,8599,1707760,00.html
[6]en.wikipedia.org/wiki/Friendster

8. DISCUSSION

In this article, we have presented the first empirical analysis of social resilience in OSN. We approached this question using a theoretical model that relates the environment of the OSN with the cascades of user departures. We showed how a generalized version of the k-core decomposition allows the empirical measurement of resilience in OSN. Previous theoretical works [5] and empirical observations [37] suggest the existence of constant cost and monotonous benefits, which lead to a stable solution that corresponds to the k-core decomposition of the social network. Among the costs that users face when using an OSN, there are time costs to adapt to the user interface and set up privacy settings [27], including the risk of revealing private information, or sharing pictures with undesired contacts [20]. The managers and owners of OSN have thus an interest in lowering this cost, usually introducing new technologies like link recommender systems or automatized friend lists [28].

We provided an empirical study of social resilience across five influential OSN, including successful ones like `Facebook` and unsuccessful ones like `Friendster`. We have shown that the hypothesis of a power-law degree distribution cannot be accepted for any of these communities, discarding the epidemic properties of complex networks as a possible explanation for large-scale cascades of user departures. Our k-core analysis overcomes this limitation, quantifying social resilience as a collective phenomenon using the CCDF of node coreness. We found that the topologies of successful sites are less resilient than the unsuccessful ones. This indicates that the environmental conditions of an OSN play a major role for its success. Thus, we conclude that the topology of the social network alone cannot explain the stories of success and failure of the studied OSN, and it is necessary to focus future empirical analysis in additional dimensions of user activity [36]. Additionally, we found very high maximum coreness numbers for most of the OSN we studied. The existence of these superconnected cores indicates that information can be spread efficiently through these OSN [23].

As a case study, we provided a detailed analysis of the changes in `Friendster` through time. We detect that the range of connections towards future nodes is much lower than the expectation from a random process. We provide an estimation of the likelihood of users to leave depending on their amount of active friends, finding that users with less active friends are more likely to leave. Not surprisingly, this likelihoood function reveals some heterogeneity among users in the decision when to leave, in line with previous research where personality traits, like extroversion, play a role in online activity [6]. Finally, we applied all our findings to `Friendster`'s collapse, fitting an approximated time series of active users through the spread of user departures predicted by the k-core decomposition.

Our analysis focused on the macroscopic resilience of OSN, but further research is necessary to better understand the individual conditions for users to leave an OSN. Clickstream datasets would allow to measure the time users spend in each social network, to quantify passive activity (viewing pictures, reading comments), and how they migrate across OSN [4]. This type of data would add an independent dimension of activity in the from of wall posts [35], picture shares and likes [7], allowing more precise validations of when users become inactive and under which situation.

Our formulation of a generalized k-core can be applied when user decisions are more complex than just staying or leaving the network, for example introducing heterogeneity of benefits or weights in the social links. For example, link weight can be estimated from implicit interactions [16], which can be incorporated to our k-core analysis through the formulation of [12]. Another open question is the role of directionality in the social network, and how to measure resilience when asymmetric relations are allowed. The benefits of users of these networks would be multidimensional, representing both the reputation of a user and the amount of information it receives from its neighborhood. The work presented here is theoretically limited to the study of monotonously increasing, convex objective functions of benefit versus active neighborhood. While empirical studies support this assumption [3, 37], it is possible to imagine a scenario where information overload decreases the net benefit of users with very large neighborhoods, creating nonlinearities where the generalized k-core is not a stable solution. We leave this questions open for further research, and the study of social resilience in other types of online communities.

9. ACKNOWLEDGEMENTS

The authors would like to thank the Internet Archive for their work in crawling and curating `Friendster`.

10. REFERENCES

[1] W. Adger. Social and ecological resilience: are they related? *Progress in Human Geography*, 24, 2000.

[2] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *WWW '07*, 2007.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks. In *KDD '06*, 2006.

[4] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *IMC '09*, 2009.

[5] K. Bhawalkar, J. Kleinberg, K. Lewi, T. Roughgarden, and A. Sharma. Preventing unraveling in social networks: the anchored k-core problem. In *ICALP'12*, 2012.

[6] C. Budak and R. Agrawal. On Participation in Group Chats on Twitter. In *WWW'13*, 2013.

[7] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *WWW '09*, 2009.

[8] H. Chun, H. Kwak, Y.-H. Eom, Y.-Y. Ahn, S. Moon, and H. Jeong. Comparison of online social relations in volume vs interaction. In *IMC '08*, 2008.

[9] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661, 2009.

[10] G. Dror, D. Pelleg, O. Rokhlenko, and I. Szpektor. Churn prediction in new users of Yahoo! answers. In *WWW '12 Companion*, 2012.

[11] S. Gaito, M. Zignani, G. P. Rossi, A. Sala, X. Zhao, H. Zheng, and B. Y. Zhao. On the bursty evolution of online social networks. In *HotSocial '12*, pages 1–8, 2012.

[12] A. Garas, D. Garcia, M. Skowron, and F. Schweitzer. Emotional persistence in online chatting communities. *Scientific Reports*, 2:402, 2012.

[13] A. Garas, F. Schweitzer, and S. Havlin. A k-shell decomposition method for weighted networks. *New Journal of Physics*, 14:083030, 2012.

[14] D. Garcia, F. Mendez, U. Serdült, and F. Schweitzer. Political polarization and popularity in online participatory media. In *PLEAD '12*, 2012.

[15] M. Giles. A world of connections - A special report on social networking. *The Economist*, Jan 28th, 2010.

[16] M. Gupte and T. Eliassi-Rad. Measuring tie strength in implicit social networks. In *WebSci '12*, 2012.

[17] A. Harkins. Network Games with Perfect Complements. Warwick University Draft, unpublished.

[18] O. Herrera and T. Znati. Modeling Churn in P2P Networks. In *ANSS '07*, 2007.

[19] B. A. Huberman. Big data deserve a bigger audience. *Nature*, 482(7385):308, 2012.

[20] M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and Privacy : It's Complicated. In *Symposium on Usable Privace and Security*, 2012.

[21] S. R. Kairam, D. J. Wang, and J. Leskovec. The life and death of online groups. In *WSDM '12*, 2012.

[22] J. Kawale, A. Pal, and J. Srivastava. Churn Prediction in MMORPGs: A Social Influence Based Approach. In *ICCSE '09*, 2009.

[23] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.

[24] J. Kleinberg. Analysis of large-scale social and information networks. *Philosophical transactions of the Royal Society A*, 371, 2013.

[25] F. Kooti, K. P. Gummadi, and W. A. Mason. The Emergence of Conventions in Online Social Networks. In *ICWSM '12*, 2012.

[26] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *WWW '08*, 2008.

[27] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings. In *IMC '11*, 2011.

[28] Y. Liu, B. Viswanath, M. Mondal, K. P. Gummadi, and A. Mislove. Simplifying friendlist management. In *WWW '12 Companion*, 2012.

[29] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *IMC '07*, 2007.

[30] L. Muchnik, S. Pei, L. C. Parra, S. D. S. Reis, J. S. Andrade, S. Havlin, and H. A. Makse. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific reports*, 3:1783, 2013.

[31] H. Park and S. Moon. Sampling Bias in User Attribute Estimation of OSNs. In *WWW '13 companion*, 2013.

[32] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3):1–4, 2002.

[33] M. P. H. Stumpf and M. A. Porter. Critical truths about power laws. *Science*, 335(6069):665–6, 2012.

[34] J. L. Toole, M. Cha, and M. C. González. Modeling the adoption of innovations in the presence of geographic and media influences. *PloS one*, 7(1):e29528, 2012.

[35] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. In *WOSN '09*, 2009.

[36] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *EuroSys '09*, page 205, 2009.

[37] S. Wu, A. Das Sarma, A. Fabrikant, S. Lattanzi, and A. Tomkins. Arrival and departure dynamics in social networks. In *WSDM '13*, 2013.

[38] E. Zheleva, H. Sharara, and L. Getoor. Co-evolution of social and affiliation networks. In *KDD '09*, 2009.